科技论文数据库作者识别号的适用性研究

史冬波[1]，邓会[2]，杨致简[3]

刘宁杰[1]，刘余秀[4]，毛宇飞[5]

[1](上海交通大学国际与公共事务学院 上海 200030)

[2](南京工业大学经济与管理学院 江苏 210000)

[3](浙江大学管理学院 浙江 310058)

[4](中国政法大学法律硕士学院 北京 100091)

[5](同济大学经济与管理学院 上海 200092)

摘要：

[目的]检验主要科技论文数据库的作者识别号的覆盖范围与准确性，并验证其能否直接用于科学学与科技政策的实证研究。

[方法]以 825 位华人科学家的发表论文为标准数据集，通过检索和收集科技论文数据库中科学家识别号及其论文信息，计算数据的覆盖率、准确性和稳健性，并运用双重差分法进行实验复现检验数据库的适用性。

[结果]第一，WOS、Scopus、AMiner 和 OpenAlex 四个数据库可检索到 90%以上的华人科学家识别符，ORCID 覆盖率不足 50%。第二，Scopus 的准确性最高为 85.2%，OpenAlex 最低仅为 51.2%；第三，直接使用数据库作者识别号的数据用于实证研究会引入不可忽视的误差。

[局限]准确集主要由青年科学家组成，学科层面未覆盖社会科学与人文科学，具有一定的局限性。

[结论]当前主要数据库的作者识别号还不能直接应用于大规模数据的实证研究，可通过建立标准化的科学家成果认证信息平台来提高中国作者姓名识别准确性。

关键词：科技论文数据库；作者识别号；姓名消歧；

分类号：G316，G353.1

DOI:

A Study on the Applicability of Author Identification Numbers in Scientific and Technical Paper Databases

Shi Dongbo[1], Deng Hui[2], Yang Zhijian[3], Liu Ningjie[1], Liu Yuxiu[4], Mao Yufei[5]

[1](School of International and Pulic Affairs,Shanghai Jiao Tong University, Shanghai, 200030, China)

[2](School of Economics and Management, Nanjing University of Technology, Jiangsu, 210000, China)

[3](School of Management, Zhejiang University, Zhejiang, 310058, China)

[4](School of Juris Masters, China University of Political Science and Law, Beijing, 100091, China)

[5](School of Economics and Management, Tongji University, Shanghai, 200092, China)

Abstract

[Purpose] To evaluate the coverage and accuracy of author identification number (author ID) of the major bibliographic databases and to assess whether they could be directly used in empirical research.

[Methods] The ground truth data set consists of articles from 825 Chinese scientists. The coverage, accuracy, and robustness of each author ID are calculated by retrieving and collecting the IDs of scientists and their respective publication information in the bibliographic databases. The validity of the author IDs for empirical research is assessed by replicating a top journal empirical article using the data collected through author IDs.

[Results] First, WOS, Scopus, AMiner, and OpenAlex can retrieve more than 90% of Chinese scientists' identifiers, while ORCID's coverage is less than 50%. Second, the accuracy of Scopus is the highest at 85.2%, and the accuracy of OpenAlex is the lowest at only 51.2%. Third, directly using the publication data collected through author IDs for empirical research will introduce non-negligible bias.

[Limitations] The ground truth data set is limited, because it is mainly composed of young scientists, and lack scientists from social sciences and humanities.

[Conclusion]At present, the author identification number of the major databases cannot be directly applied to the empirical research of large-scale data. A standardized information platform for scientists' publications is needed to overcome the author-name disambiguation problem.

Keywords：Bibliographic databases; Author identification number; Author-name disambiguation;

1 引言

当今世界百年未有之大变局加速演进，世界经济陷入下行周期，各主要大国围绕科技制高点的竞争空前激化，科技创新成为国际战略博弈的主要战场。科技竞争的决定性力量在于人才，实施人才强国战略已经成为党和国家一项重大而紧迫的任务。习近平总书记在中央人才工作会议上指出，尽管“我国已经拥有一支规模宏大、素质优良、结构不断优化、作用日益突出的人才队伍”，但是“人才发展体制机制改革“破”得不够、“立”得也不够，既有中国特色又有国际竞争比较优势的人才发展体制机制还没真正建立”[1]。十八大以来，我国科学研究取得新的历史性成就，我国高质量论文首次跃居世界第一[2]，我国正处于从量到质、从追赶到引领的关键节点。在研发投入持续增长与高等教育长足进步背景下，建设符合科学研究规律、支持原始创新的人才体制机制，是建设科技强国的关键。其中，包括薪酬设计[3]、人才评价在内的激励制度是人才发展体制机制的基础[4]，关乎我国科技资源投入到产出的转化效率，亟需长期深入研究。

人才机制体制的研究离不开科学学理论与实证研究的支撑。人才评价、人才计划还是激励制度的改革，都需要建立在精准的政策评估的基础上。这要求研究单位和研

究数据从以往的地区与单位层面，精细化到科学个人与团队层面。其中，科技论文数据是不可或缺的基础数据，Web of Science、Scopus 等数据库常被用来研究科学家的评价[5,6]、流动[7,8]与激励[9]等问题。但是，大量科学家共享了同样的姓氏与名字（或名字首字母），致使将数据库中姓名相同的作者区分为现实中不同的科学家（作者姓名消歧）成为一个较大挑战，这一现象在华人群体中尤其严重[10]。不解决这个问题，就无法准确进行科学家层面的实证研究，理论研究和政策研究更无从谈起。

因此，本文使用 Shi 等[11]搜集的 825 位华人科学家发表论文标准数据集来检验主要科技论文数据库 Web of Science、Scopus、OpenAlex、ORCID 以及 AMiner 的作者识别号的覆盖范围与准确性，并通过复现实验检验数据库识别号能否直接用于实证研究。本文的章节安排如下，第二部分对相关研究进行梳理，第三部分介绍研究数据与方法，第四部分介绍研究结果，最后进行总结与讨论。

2 相关研究

当前作者姓名消歧的方式有两种：通过算法自动生成和作者自我汇报（认领）。前者的覆盖范围更全，后者的准确性更高。作者姓名消歧算法使用分析型或模式识别型的算法将作者姓名相似的论文进行聚类，自动生成作者识别号[12,13]。其中最著名的是由 Torvik 和 Smalheiser 于 2009 年开发的针对 MEDLINE 数据库的作者姓名消歧算法，作者的算法和消歧数据最终被整合进入 PubMed 数据库[14]，为诸如科学家合作[15]、科学研究方向选择[16]、性别问题[17]、同行评议[18]等科学学与科学经济研究议题奠定了基础。遗憾的是，Torvik 和 Smalheiser 的数据仅支持医学与生命科学领域的研究，无法应用在更广泛的学科上[14]。OpenAlex 使用机器学习算法将其所有论文作者进行了姓名消歧处理，并开源了算法源代码和数据 [1]，这为科学学研究注入新的动力。此外，大多数其他消歧算法的作者并没有提供开源的算法与数据，复现算法所需要的算力和资源也往往超过了科学学研究人员的能力。

为了解决数据库姓名歧义的问题，各主要科技论文数据库运营商与其他非营利组织选择了另一条技术路线。2008 年，Web of Science 数据库（WOS）推出了身份唯一识别符 ResearcherID，科学家可以注册 ResearcherID，自行认领 Web of Science 数据库内的论文。2012 年，非盈利组织 Open Researcher and Contributor Identifier（ORCID）发布用户标识符，作者可以注册 ORCID，并在其平台维护个人的学习与工作履历，以及论文发表记录。如今，很多国际期刊要求作者在提交初稿时同时指定其 ORCID[19]。Scopus 数据库的 Scopus Author Identifier（Scopus AuthorID）则综合了自动生成算法与科学家自主反馈的方式 [2]。

[1] 算法的说明参考 https://docs.openalex.org/api-entities/authors/author-disambiguation；源代码位于 https://github.com/ourresearch/openalex-name-disambiguation/tree/main。
[2] Web of Science 目前也采用了自动生成算法与科学家自主认领相结合的方式。

数据库的作者识别号为以科学家个人或团队为研究单位的科学学研究提供了新的高质量研究数据。例如，Moed 等使用 Scopus AuthorID 来研究移民科学家[20]，Khurana 和 Sharma 联合使用 Researcher ID, AuthorID 和 ORCID 来研究 h 指数如何用于科学家的评价[21]。相关数据近些年开始被应用于中国科学家的研究，如 Zhao 等使用 ORCID 的数据证实海归科学家并没有表现出比本土科学家更强的学术发表能力[22]，这一结论与学术界的认知相悖 [3]。

科技论文数据库作者识别号的准确性与覆盖范围直接影响了使用这些数据的实证论文的信度与效度。使用不准确的数据得出的结论可能是具有误导性的，使用准确但是覆盖范围不全的数据得出的结论往往缺乏代表性。因此，必须检验科技论文数据库作者识别号的适用性。Aman 使用 193 名德国莱布尼兹奖获得者的数据证实了 Scopus AuthorID 的查全率和精准度分别高达 97%和 100%[23][4]，并且证实可以用 Scopus AuthorID 来追踪科学家的跨国流动。Kawashima1 和 Tomizawa 使用日本科学资助数据库 KAKEN 证实 Scopus AuthorID 的查全率和精准度分别为 98%和 99%[24][5]。Boudry 和 Durand-Barthez 则发现 ORCID 与 ResearcherID 对一组法国科学的覆盖率均不足 20%，且大量 ID 没有涵盖完整的发表记录[25]。可见，科技论文数据库作者识别号的准确性与覆盖范围针对不同的群体差异显著。特别是，当前的研究中没有针对华人科学家群体的检验，这便限制了相关作者识别符在我国科学学与科技政策研究中的应用。

3 数据与方法

3.1 标准数据集

本文使用 Shi 等[11]搜集的 825 位华人科学家发表论文标准数据集（表 1）。该数据集涵盖了一批于 1997 年至 2014 年之间获得博士学位的华人科学家，平均毕业年份为 2007 年。其中，14%为女性，18%在中国大陆取得博士学位，65%在美国获得博士学位。截止 2019 年，49%的科学家在中国大陆的学术机构工作，42%在美国的学术机构工作，其他科学家主要在欧洲、日本与中国香港地区工作。该数据集涵盖了所有自然科学的领域，工程与材料科学和医学领域的科学家最多，分别占到 22%与 21%；地球科学领域的科学家最少，但占到了 9%。因此，该数据集作为标准数据集，具有一定代表性。

[3] 后文中将看到，这一结论很有可能是由于 ORCID 数据的缺失造成的。

[4] 此处需要注意的是，作者并没有收集到完整的科学家发表清单，作者定义的查全率是主要 Scopus AuthorID 覆盖论文占到所有 AuthorID 论文的比重，因此作者可能高估了查全率。

[5] 作者估计的是科学资助级别的查全率和精准度，资助项目往往只能代表一个科学家 3-5 年的发表记录，在这样的时间维度上，查全率和精准度都很有可能被高估。比如，一个科学家可能在不同的职业生涯阶段分别主持了不同的研究项目，算法较容易将不同阶段的同一个科学家识别成为不同的作者，从而生成不同的 ID，从项目级别来看，这样的 ID 非常准确，但是从科学家的几倍来看，这样的 ID 每一个都不够准确。

Shi 等[11]从科学家的个人主页（41%）、谷歌学术（39%）、Researchgate（12%）等来源收集了这些科学家从博士毕业开始至 2019 年的发表在 SCI/SSCI 索引期刊的所有论文发表记录[6]（见表 2）。数据集中科学家平均每人发表论文 56 篇[7]。

表 1 标准数据集特征

Table1 Characteristics of the standard dataset

变量	样本量	均值	标准差	最小值	最大值
博士毕业年份	824	2007	2.69	1997	2014
女性	825	0.14	0.35	0	1
在中国大陆获得博士学位	825	0.18	0.38	0	1
在美国获得博士学位	825	0.65	0.48	0	1
在中国大陆工作（2019 年）	825	0.49	0.50	0	1
在美国工作（2019 年）	824	0.42	0.49	0	1
数学与物理	825	0.18	0.39	0	1
化学	825	0.15	0.36	0	1
信息科学	825	0.14	0.35	0	1
生命科学	825	0.21	0.41	0	1
工程与材料科学	825	0.22	0.42	0	1
地球科学	825	0.09	0.29	0	1
论文数量	825	56.03	73.34	1	1174
大陆科学家的论文数量					
海外华人科学家的论文数量					

表 2 标准数据集数据来源

Table2 Data sources of the standard dataset

数据来源	个人认证	无个人认证
个人简历与主页	341	
谷歌学术	314	5
Rearchgate	99	
ORCID	31	
Publons	12	13
PubMed/ INSPIRE/Linkin	10	

注：谷歌学术与 Publons 会显示该账号是否经过科学家个人认证。

3.2 科技论文数据库作者识别号与发表论文

本文通过标准数据集中科学家的工作履历以及研究领域从科技论文数据库检索科学家对应的作者识别号。本文选择科学学与科技政策研究中最常用的四个科技论文索引数据库，Web of Science，Scopus，OpenAlex 与 AMiner，前三个数据库提供了作者个人识别号（AuthorID），AMiner 则提供了包含论文列表的科学家个人主页。

ORCID（Open Researcher and Contributor Identifier，开放研究者与贡献者身份识别码）是由非营利性组织 ORCID 于 2012 年 10 月 16 日推出并发布的用户标识符。通过给每位注册的科学家分配唯一的 16 位数标识符，为研究者提供唯一的身份标识。科学

[6] 其中，60 位科学家的数据集存在若干年份缺失。对于这些科学家，在后续计算时，缺失年份的论文数据统一进行了删除。特别地，如果使用 765 位拥有完整论文发表数据集的科学家论文集作为标准数据集，本文的研究结论不会发生改变。

[7] 数据集中包括三位高能物理领域的科学家，分别发表了 850 篇、919 篇和 1174 篇论文，从标准数据集中删除这三位科学家不会改变本文的研究结论。

家可以将在 ORCID 平台中关联自己发表在 WOS 与 Scopus 中的论文[26]。2012 年 ORCID 系统整合进入 WOS ResearcherID。特别需要注意，论文准确集中包含 ORCID 来源数据中没有谷歌学术或者 Researchgate 的学者，且 ORCID 发表记录完整的科学家。此处要检验的是 ORCID 作为单一数据源的效果，与标准集不同。

Web of Science（WOS）创建于 1964 年，覆盖了自然科学、社会科学、艺术和人文学科等全球范围内的学术期刊、会议论文和引用数据，其科学引文索引（SCI）和社会科学引文索引（SSCI）数据集是科学学与科技政策研究的权威数据集。截止到 2023 年 12 月，SCI 共收录超过 9,500 本杂志和 6,100 万篇论文，SSCI 共收录超过 3,500 本杂志和 1,000 万篇论文。2008 年起，Web of Science 推出身份唯一识别符 ResearcherID。一开始，ResearcherID 系统要求用户自行注册，注册后，可以将自身的 ResearcherID 与 Web of Science 中的论文进行连接。之后，ResearcherID 引入身份自动生成算法将没有作者认领的论文分类生成作者识别符[27]。

Scopus 是 Elsevier 于 2004 年推出的摘要和引文数据库，完整数据库可以追溯到 1966 年，包括生命科学、社会科学、自然科学和医学领域。Scopus 数据库使用数据库中记录的作者及其出版物的信息，如所属单位、学科领域、文章标题、引用和合著者，基于先进算法为每位作者分配一个唯一的标识符，即 Scopus Author Identifier，可以自动区分同名作者以及匹配作者姓名的变化[20]。

AMiner 于 2006 年 3 月推出，是新一代科技情报分析与挖掘平台，由清华大学计算机科学与技术系教授唐杰率领团队建立，聚合了全球各个国家和地区的学者画像、机构画像、期刊画像等数据，覆盖各个学科领域包括自然科学、社会科学、人文科学等[28]。AMiner 从分布式网络中提取和整合学术数据，为每位研究人员创建基于语义的个人资料，使用生成概率模型对论文、作者和发表地点等主题方面进行建模，分析和发现研究人员社交网络中的有趣模式，以及基于建模结果实现诸如专业知识搜索和关联搜索的若干搜索服务；为研究人员提供了一个档案数据集[29]。

OpenAlex 是 OurResearch 在 2022 年 1 月推出的是一个免费开放的全球学术研究数据库，收录了各学科领域的开放获取期刊和研究成果[30]。OpenAlex 由研究成果、作者、机构、场地和概念五种类型的实体组成，继承了微软学术（Microsoft Academic Graph）数据，并通过机器学习算法对所有作者进行了姓名消歧[31]。

本文在以上五个数据库中对标准数据集中的科学家进行检索，并记录下对应的科学家识别号及其论文信息。ORCID、WOS、Scopus 与 AMiner 四个数据库具备网页检索功能，具体检索步骤，如图 1 所示。首先在数据库检索界面输入科学家信息进行检索，如姓名、工作机构等；其次根据标准数据集中科学家的教育背景、工作经历、研究领域和起始时间等关键信息，从检索结果中选取匹配的且有发表记录的识别号，如果存在多个与相关信息匹配的科学家识别号，则记录下前三位最为匹配的识别号

（AMiner 则记录下当前科学家所属网页地址）；最后收集科学家论文信息，包括发表论文 doi 号、标题、期刊、发表年份、入藏号、作者等。OpenAlex 提供的是 PostgreSQL 数据[8]，本文通过全名检索在 authors 表中匹配得到作者 id 信息，并利用该 id 信息与 work 表格中 author_id 列匹配进而得到其论文 id（work_id），以及每篇论文中作者的机构信息（raw_affiliation_string）。进一步从数据库中得到论文相关信息，包括发表论文 doi 号、标题、期刊、发表年份、入藏号、作者等。后续分析中，我们将使用准确性最高的 ID 作为评判标准。本文的附件以科学家 chen jianing 为例描述了在不同数据库中的检索流程以及科学家标识号和论文的收集记录。

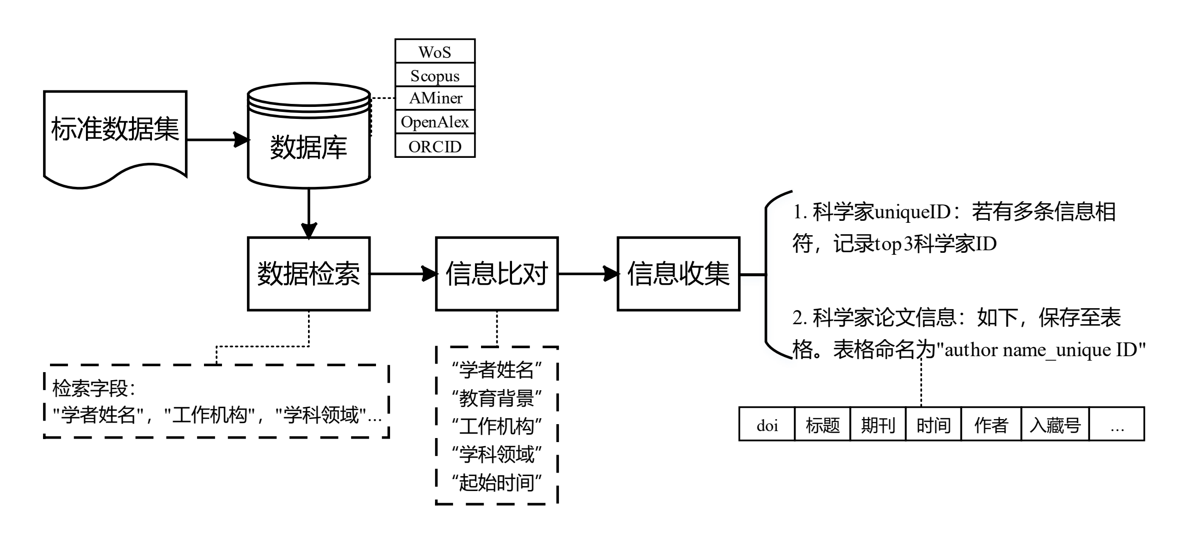


图 1 数据库检索基本流程图
Fig.1 Basic Flow Chart of Database Retrieval

3.3 数据库论文集与标准数据集之间的连接

本文采用 WOS 数据库的论文入藏号来连接 ORCID 与 WOS 论文集与标准数据集。我们找到了 402 名科学家的 ORCID，总计 24,181 篇论文；我们为 777 名科学家检索到 1,115 个 WOS ResearcherID，总计 45,485 篇论文。

对于 Scopus、AMiner 和 OpenAlex，分别按照下面步骤连接数据库论文集与标准数据集：

第一步，限制数据库论文集范围。由于标准数据集仅收录了科学家发表在 SCI 与 SSCI 索引杂志的论文，本文首先根据 Journal Citation Report 中每年收录杂志的清单，将数据库论文集限制在 SCI 与 SSCI 收录论文范围内。同时根据标准数据集对每个科学家的覆盖年份，将数据库论文集现在相同年份发表的论文。如表 3 所示，Scopus、AMiner 以及 OpenAlex 的数据被 SCI/SSCI 覆盖的比例分别为 72.9%、67.1%以及 61.9%。

[8] OpenAlex 数据的具体获取方式参见 https://docs.openalex.org/download-all-data/download-to-your-machine

第二步，通过数字对象唯一标识符（DOI）连接论文。在前一步骤基础上，如果论文的 DOI 相同，直接连接。Scopus、AMiner 和 OpenAlex 三个数据库的连接比例分别为 77.2%、65.5%以及 4.7%。由于 OpenAlex 涉及到的备选 ID 与论文数量比其他数据库高出两个量级，匹配到表述数据集论文的比例要显著更小。

第三步，通过发表期刊、发表年份与标题精确匹配。对于 DOI 信息缺失的论文（数据库论文集与标准数据集其一缺失），本文通过发表期刊、发表年份与标题精确匹配。三个数据库在这一步的匹配率分别为 5.1%、4.2%以及 0.2%。

第四步，通过标题模糊匹配，人工检查。对于发表期刊和年份精确匹配，但是标题无法精确匹配的论文对，本文计算两篇论文之间的标题相似度（定义为去除符号后的论文标题重合单词数量占论文单词总数的比例），然后对相似度超过 80%的论文进行人工比对确认是否为同一篇论文 [9]。这一步骤中匹配到 0.9%的 Scopus 论文和 0.7%的 AMiner 论文。OpenAlex 涉及的论文数量过于庞大，只能略去人工校对这一步。因此，OpenAlex 的准确性可能会被低估 1%左右。但是，后文中我们将会看到，这一比例对数据库最终的准确性评价的影响可以忽略。

表 3 数据库论文集与标准数据集匹配过程

Table3 Matching process between database paper collection and standard dataset

数据库	ID 数量	ID 覆盖人数	论文数量	SCI/SSCI 论文数量	DOI 匹配论文数量	期刊/年份/标题精准匹配论文数量	标题模糊匹配与人工校对论文数量	准确集论文对应作者的数量
ORCID	402	402	22,778	22,778	-	-	-	24,181
WOS	1,151	777	58,073	58,073	-	-	-	45,485
Scopus	847	813	63,702	46,430 (72.9%)	35,833 (77.2%)	2,390 (5.1%)	433 (0.9%)	46,330
AMiner	891	757	69,846	46,878 (67.1%)	30,717 (65.5%)	1,951 (4.2%)	351 (0.7%)	45,163
OpenAlex	16,870	798	1,157,866	717,032 (61.9%)	33,851 (4.7%)	1,387 (0.2%)	-	42,580

通过四个步骤，我们可筛选出每一个数据库论文集与标准数据集之间的交集，用于评价数据库作者识别号的效果。

3.4 判断指标

本文使用以下指标评估数据库作者识别号的覆盖率、准确性与稳健性。对于检索得到多个作者识别号的科学家，本文将其平均指标作为最终指标。

覆盖率（CV）：该数据库中满足检索条件可获取作者识别号的科学家人数占总人数的比例。该指标决定了数据库的适用范围。

B3 精准度（B3 precision，BP）、B3 查全率（B3 recall，BR）与 B3F1 分数（B3 F1-score，BF1），定义如下：

[9] Openalex 涉及到的论文数量过大，省去这一步。不过，由于这一步实际匹配的论文数量很少，因此对于最终结果的影响很小。

$$BP = \frac{1}{N} \times \sum_i \frac{|D_i \cap G_i|}{|G_i|} \quad (1)$$

$$BR = \frac{1}{N} \times \sum_i \frac{|D_i \cap G_i|}{|G_i|} \quad (2)$$

$$BF1 = \frac{2 \times BP \times BR}{BP + BR} \quad (3)$$

其中，G_i表示科学家 i 的标准论文数据集，D_i表示科学家在数据库对应作者识别符下论文数据集，N 表示数据库中检索得到的科学家人数，$|D_i|$表示数据集中的元素数量。B3 准确性指标是文献中常用的衡量算法准确性的指标[32]。BP 描述的是数据库中识别出来的科学家论文有多少比例确实是科学家发表的，BC 描述的是科学家实际发表的论文有多少比例确实被数据库识别出来。可见 BP 与 BC 存在某种平衡关系，例如，一个精准度搞得算法可能会遗漏掉更多的论文，因此，本文使用其调和平均数来表示 BP 与 BC 的平均表现。

B3 准确性指标衡量的是每个科学家识别号的准确性的平均值。为了衡量数据库识别号的稳健性，本文引入精准度与查全率的标准差，定义如下：

$$SDP = sd\left(\frac{|D_i \cap G_i|}{|G_i|}\right) \quad (4)$$

$$SDR = sd\left(\frac{|D_i \cap G_i|}{|G_i|}\right) \quad (5)$$

3.5 实证实验

为了进一步验证作者识别号能否用于科学学与科技政策的实证研究，本文将不同数据库中识别出的数据集复现 Shi 等[11]的研究，检验不同数据库能够得出与标准数据集一致的研究结论。Shi 等[11]的研究问题为青年华人科学家回国后职业生涯（相比于其在海外学术界工作的同学而言）能否更加成功？文章作者使用了标准的双重差分方法，以论文发表数量为因变量，以归国科学家与归国前后的指标变量的交乘项为核心自变量，控制了个人与年份的固定效应。回归方程为：

$Y_{i,t} = \alpha + \beta\ PostReturn_{i,t} * Treat_i + PostReturn_{i,t} + \gamma_i + \eta_t + \varepsilon_{i,t}$。

其中，$Y_{i,t}$表示科学家 i 在 t 年的论文发表数量，$PostReturn_{i,t}$表示科学家 i 在 t 年是否回国，$Treat_i$表示科学家是否为归国科学家。同时，论文采用了匹配的策略，只有年龄与学习经历相似，科研能力接近的科学家才能最终进入回归。

4 实证结果

4.1 覆盖率与准确性

如表 4 所示，WOS、Scopus、AMiner 以及 OpenAlex 四个数据库的识别号覆盖率均达到 91%以上，其中 Scopus 的覆盖率最高为 98.5%，OpenAlex 第二高为 96.7%。值

得注意的是，ORCID 的覆盖率仅仅为 48.7%，远低于作者的预期。这一结果虽然好于 Boudry 和 Durand-Barthez[25]的发现，但仍然意味着有超过一半的华人科学家没有注册 ORCID 或者没有在 ORCID 中维护个人信息，以至于无法检索得到其 ORCID。注意到标注数据集中的科学家已经是青年科学家群体，如果考虑更加资深的华人科学家，ORCID 的覆盖率可能还会更低。因此，从覆盖率的角度来看，WOS、Scopus、AMiner 以及 OpenAlex 四个数据库中可以找到绝大多数的华人科学家识别符，可以用于实证研究；但是 ORCID 的覆盖率不足一半，用于实证研究可能会带来不可忽略的偏差。

表 4 数据库识别号的覆盖率与准确性

Table4 Coverage and accuracy of database identification numbers

数据库	人均论文数量	标准集人均论文数	CV	BP	BR	BF1	SDP	SDR
ORCID	56.66	59.76	0.487	0.826	0.738	0.780	0.214	0.366
WOS	64.09	57.38	0.942	0.645	0.728	0.684	0.311	0.356
Scopus	56.49	54.03	0.985	0.831	0.874	0.852	0.158	0.180
AMiner	57.52	58.47	0.918	0.736	0.724	0.730	0.224	0.268
OpenAlex	168.50	56.33	0.967	0.397	0.724	0.512	0.254	0.299

不同数据库识别号的精准度差异显著。Scopus 的精准度最高为 83.1%，但这一数字远低于 Aman[23]和 Boudry 和 Durand-Barthez[25]汇报的结果，说明 Scopus 的算法在华人科学家群体中的表现低于其在其他族群科学家群体中的表现。令人意外的是，ORCID 的精准度虽然比 WOS、AMiner 和 OpenAlex 更高，但也仅仅为 82.6%，并没有达到预想中的接近 100%（理论上 ORCID 的数据是作者个人维护的，因此应该非常精准）。本文认为两个原因共同导致了这一现象：第一，检索到了错误的 ORCID 导致精准度为零，这部分占到所有识别号的 2.7%；第二，ORCID 允许作者使用第三方平台（如 Scopus、Crossref）来管理其个人数据。当作者将第三方平台的识别号与 ORCID 连接后，ORCID 会自动将相关平台的数据导入至 ORCID 中，从而降低了精准度。WOS 和 AMiner 的精准度分别为 63.5%与 73.6%，而 OpenAlex 的精准度仅仅为 39.7%，这意味着这三个数据库给科学家分配了非本人发表的论文。

不同数据库识别号的查全率相对接近。Scopus 的查全率最高为 87.4%，其他四个数据库的查全率在 72.4%至 73.8%之间。WOS、Scopus 与 OpenAlex 的查全率高于精准度，这导致其作者识别号会高估科学家的发表数量。其中，OpenAlex 的偏差最大，高估接近了 200%。相反，ORCID 与 AMiner 的识别号则会低估科学家的论文数量，ORCID 人均低估了 3.1 篇论文。

综合来看，Scopus 的准确性最高，F1-分数达到 85.2%，比其他数据库至少高出 7%，这可能得益于 Scopus 团队对作者姓名消歧算法的重视与持续改进，以及对华人科学家群体数据集的关注。此外，Scopus 数据库的稳健性显著高于其他四个数据库。如图 2 所示，Scopus 的综合表现最好。OpenAlex 的准确性最差，F1-分数仅仅为 51.2%，造成这一结果的原因可能是：第一，OpenAlex 没有引入科学家个人认证与校对的机制，第二，OpenAlex 没有使用高质量的华人科学家数据集来训练其算法。

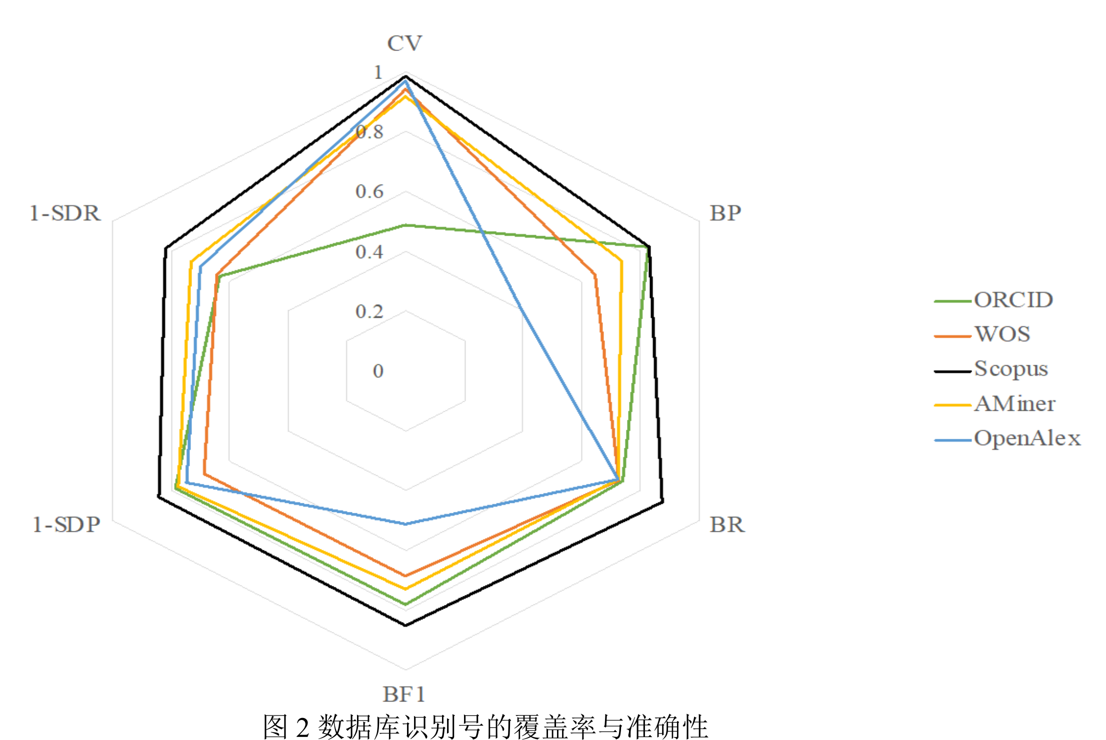


图 2 数据库识别号的覆盖率与准确性
Fig.2 Coverage and accuracy of database identification numbers

4.2 异质性

作者姓名消歧工作实际上是将某个作者的论文（准确论文）从一组作者姓名一致的论文（备选论文）中识别出来。备选论文的信息与准确论文的信息越接近，作者姓名消歧的挑战越大。当科学家工作的单位与领域中同名人数更多时，筛选出其准确论文的难度会更大。而工作单位（包括地区）与领域往往是实证研究中的重要变量，数据的偏差将直接扭曲研究结论。

本文将每个科学家的工作经历分为大陆与海外两部分，分别考察数据库作者识别号针对不同地区华人科学家的准确性。如表 5 所示，不同于前文的猜测，大部分数据库的作者识别号（除 ORCID 之外）反而对华人学者在大陆工作期间发表的论文准确度更高。另外，除 AMiner 之外，各数据库都会高估科学家的论文数量，同时对于大陆工作期间的科学家高估的数量更多。

表 5 数据库识别号对不同地区华人科学家的准确性
Table5 Accuracy of database identification numbers for Chinese scientists in different regions

数据库	地区	人均论文数量	标准集人均论文数	BP	BR	BF1
ORCID	海外	39.17	38.96	0.819	0.817	0.818
	大陆	42.47	46.99	0.845	0.764	0.803
WOS	海外	45.50	33.14	0.590	0.783	0.673
	大陆	62.10	49.48	0.631	0.805	0.708
Scopus	海外	32.92	31.93	0.826	0.859	0.842
	大陆	47.72	44.52	0.837	0.884	0.860
AMiner	海外	32.26	32.62	0.717	0.695	0.706
	大陆	49.36	49.01	0.765	0.786	0.776
OpenAlex	海外	108.32	32.78	0.372	0.713	0.489
	大陆	127.80	48.89	0.438	0.721	0.545

本文进一步考察了数据库识别号准确性（F1 分数）的学科差异。本文将科学家分为化学、地球与环境科学、工程与材料科学、信息科学、生命科学以及数理科学六个领域。如图 3 所示，各数据库（OpenAlex 外）在信息科学领域的准确性都远低于其他科学领域。Scopus 准确性的学科差异较小，且均高于其他数据库的最高水平。

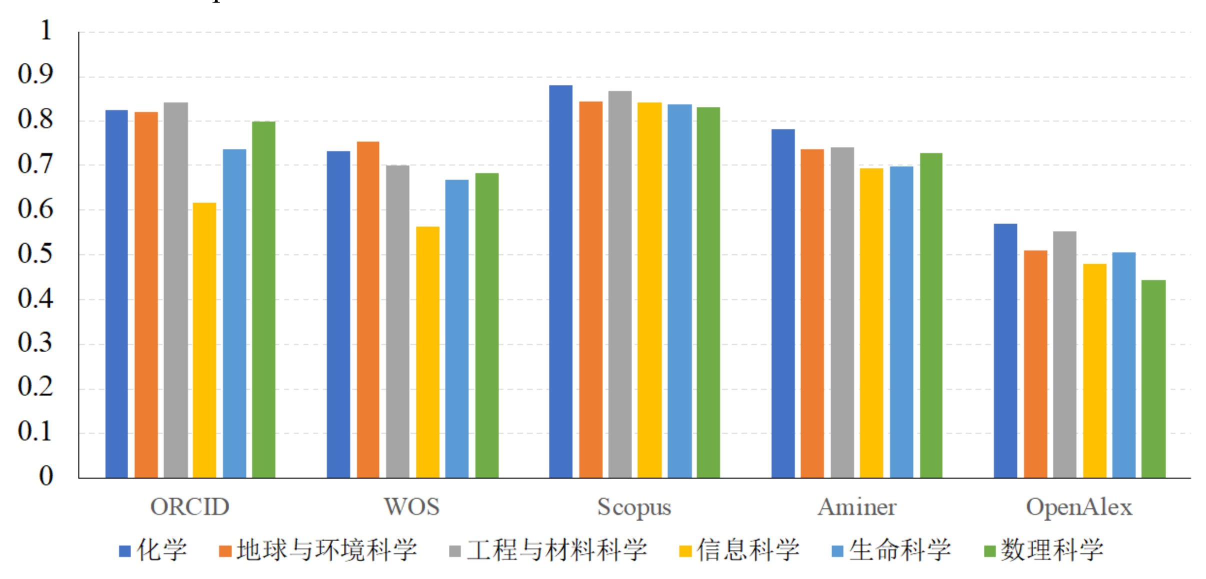


图 3 数据库识别号准确性的学科差异

Fig.3 Disciplinary differences in the accuracy of database identification numbers

最后，本文将各数据库识别号准确性（F1 分数）与科学家的个人特征进行回归。如表 6 所示，ORCID、WOS 与 Scopus 数据库对于越年轻的科学家准确性越搞，同时，ORCID 与 AMiner 对于在大陆工作的科学准确率更高，但是高出的幅度有限。而 OpenAlex 数据库对于女性科学家准确性低于男性科学家，这意味着使用 OpenAlex 可能会错误估计科学家科研效率的性别差异。

表 6 识别号准确性与科学家个人特征

Table6 Accuracy of identification number and personal characteristics of scientists

	(1) ORCID	(2) WOS	(3) Scopus	(4) Aminer	(5) OpenAlex
女性	-0.029	-0.021	-0.001	-0.023	-0.065**
	(0.051)	(0.035)	(0.017)	(0.027)	(0.027)
毕业年份	0.012*	0.013***	0.006***	0.005	0.004
	(0.006)	(0.005)	(0.002)	(0.003)	(0.004)
2019 年在大陆工作	0.063*	0.033	-0.009	0.067***	0.008
	(0.034)	(0.026)	(0.013)	(0.020)	(0.022)
其他控制变量					
学科	是	是	是	是	是
博士学位国家	是	是	是	是	是
样本量	401	774	809	753	794
对数似然	0.084	0.059	0.027	0.025	0.028

注：（1）-（5）中的模型设定为一般线性回归，以识别号的 F1 分数为因变量；样本量与表 4 的差异源自变量缺失；标准误差在括号内；显著性水平：* p<0.1,** p<0.05, *** p<0.01。

4.3 实证研究复现结果

表 7 中汇报了使用不同数据库中识别出的数据集复现 Shi 等[11]的研究的结果。其中，（1）列中汇报了 Shi 等[11]的标准数据集的结果，（2）-（6）列使用各数据库的结果。如表 7 所示，标准数据集的系数估计为 0.210（p<0.01）。使用 ORCID 与 OpenAlex 进行同样的估计得到的系数均不显著，这可能是由于 ORCID 的样本量过少，而 OpenAlex 的准确性较低。尽管使用 WOS、Scopus、AMiner 数据估计模型可以得到正显著的系数，但是，这三个模型都会高估实际系数，高估的幅度分别为 55%、99% 与 85%，高估的幅度是不可以忽略的。尤其是，尽管 Scopus 数据库在覆盖率、准确性与稳健性方面均好于其他数据库，但是其高估的幅度高达 99%，极大限制了该数据库的适用性。因此，本文认为，基于本文的实践，目前的五个数据库识别号还不能直接应用于实证研究。

表 7 复现实证研究的结果

Table7 Results of the reproduction of the empirical study

	(1) GT	(2) ORCID	(3) WOS	(4) Scopus	(5) AMiner	(6) OpenAlex
海归*回国后	0.210***	0.289	0.326***	0.418***	0.388***	0.094
	(0.076)	(0.152)	(0.094)	(0.103)	(0.096)	(0.071)
样本量	4,191	688	2,530	3,019	2,329	3,092
对数似然	-8,276	-1,363	-5,542	-6,781	-5,213	-7,966

注：（1）-（6）的模型中包括了个体固定效应与年份固定效应，个体层面的聚类标准误差在括号内；显著性水平：* p<0.1,** p<0.05, *** p<0.01。

5 结论与讨论

本文使用 Shi 等[11]搜集的 825 位华人科学家发表论文标准数据集来检验科技论文数据库 Web of Science、Scopus、OpenAlex、ORCID 以及 AMiner 的作者识别号的覆盖范围与准确性。本文发现，数据库识别号的准确性差别较大，处于 51.2%至 85.2%之间，Scopus 的准确性最高，OpenAlex 的准确性最低。其中，在 WOS、Scopus、AMiner 以及 OpenAlex 四个数据库中可以找到绝大多数的华人科学家识别符，可以用于实证研究；但是 ORCID 的覆盖率不足一半，用于实证研究可能会带来不可忽略的偏差。本文进一步揭示数据库识别号的准确性受到科学家工作地区以及学科的影响，最后通过复现实验证实目前的数据库识别号还不能直接应用于实证研究。

那么应该如何使用科学家论文数据来进行实证研究？这一问题的答案与具体研究样本与分析单位密切相关。当分析单位具体到个体自然人且样本量不大时，本文建议研究人员收集科学家的个人完整履历，并利用刘玮辰[33]与 Shi[11]等开发的基于科学家职业经历和引文网络的姓名消歧算法，这一算法的准确度均显著高于数据库的作者识别号，运算效率高，且得到了国际顶尖期刊的认可。当进行大规模数据分析时，前述算法并不适用，本文建议研究人员首先使用小规模准确集数据对数据库的作者识别号

进行检验，并在文中汇报研究结果的稳健性。此外，本文呼吁国内相关机构建立标准化的科学家成果认证信息平台，为每一个国内工作的科学家分配唯一识别号，并在政策层面激励每个科学家主动维护成果信息。这不仅有助于从源头上解决我国科学家的论文作者姓名歧义问题，还可以为改进姓名消歧算法积累宝贵的训练数据集。

参考文献

[1] 习近平. 深入实施新时代人才强国战略加快建设世界重要人才中心和创新高地[J].求是,2021(24):4-15.(Xi Jinping. Deepening the Implementation of the Strategy for Strengthening the Nation with Talents in the New Era to Accelerate the Construction of a World-Class Talent Hub and Innovation Highland[J]. Qiushi,2021(24):4-15.)

[2] Woolston C. Nature Index Annual Tables 2023: China tops natural-science table[J]. Nature, 2023.

[3] 曹艺凡,盛创新,童锋. 粤港澳大湾区引进外籍战略科学家的问题与对策[J].科技管理研究,2023,43(14):78-84.(Cao Yifan, Sheng Chuangxin, Tong Feng. Problems and Countermeasures of Introducing Foreign Strategic Scientists to the Guangdong-Hong Kong-Macao Greater Bay Area[J]. Science and Technology Management Research, 2023,43(14):78-84.)

[4] 魏海勇,李祖超. 知识型人才激励模型的建立与应用:基于成就需要理论的视角[J]. 科技进步与对策, 2008(6): 169-171.(Wei Haiyong, Li Zuchao. The Establishment and Application of Knowledge-based Talent Incentive Model: Based on the Perspective of Achievement Needs Theory[J]. Science & Technology Progress and Policy, 2008(6): 169-171.)

[5] 王甲旬,邱均平.我国化学领域青年科技人才论文产出分析——以前 5 批青年千人为例[J].现代情报,2019,39(02):8-16. (Wang Jiaxun, Qiu Junping. An Analysis of Chemistry Academic Paper: A Case Study of the First Five Batches of “Recruitment Program for Globa al Young Experts”[J]. Journal of Modern Information, 2019, 39(2):8-16.)

[6] 张丽华,吉璐,陈鑫. 科研人员职业生涯学术表现的差异性研究[J].科研管理,2021,42(5):182-190.(Zhang Lihua, JiLu, Chen Xin. A study of the difference of researchers' academic performance during their professional career[J]. Science Research Management, 2021,42(5):182-190.)

[7] 陈凯华,杨一帆,陈光,张汝昊.全球科研人员流动规律与不同层次人才的差异化研究——基于 Scopus 百年论文数据的研究[J].科学学与科学技术管理,2023,44(04):3-20.(CHEN Kaihua, YANG Yifan, CHEN Guang, ZHANG Ruhao. Research on the Characteristics of Global Talent Flow at Different Levels: Based on Paper Data of a Hundred-year from Scopus[J]. Science of Science and Management of S.& T, 2023,44(04):3-20.)

[8] 魏立才,黄祎.学术流动对回国青年理工科人才科研生产力的影响研究——基于 Web of Science 论文分析[J].高等工程教育研究,2020,(01):67-73.(Wei Licai, Huang Yi. On the Influence of Academic Mobility on the Scientific Research Productivity of Returned Young Science & Engineering Talents[J]. Research in High er Education of Engineering, 2020,(01):67-73.)

[9] 滕广青,吕晶,江瑶,庹锐,彭洁.基于 STM 的科研资助对研究主题影响研究[J].现代情报,2022,42(05):58-68.(Teng Guangqing, Lyu Jing, Jiang Yao, Tuo Rui, Peng Jie. Impact of Research Funding on Research Topics Based on STM. [J]. Journal of Modern Information, 2022,42(05):58-68.)

[10] Kim J, Kim J, Kim J. Effect of Chinese characters on machine learning for Chinese author name disambiguation: A counterfactual evaluation[J]. Journal of Information Science, 2023, 49(3): 711-725.

[11] Shi D, Liu W, Wang Y. Has China's Young Thousand Talents program been successful in recruiting and nurturing top-caliber scientists?[J]. Science, 2023, 379(6627): 62-65.

[12] Elliott S. Survey of author name disambiguation: 2004 to 2010[J]. Library Philosophy and Practice, 2010, 473: 1-11.

[13] Ferreira A A, Gonçalves M A, Laender A H F. A brief survey of automatic methods for author name disambiguation[J]. Acm Sigmod Record, 2012, 41(2): 15-26.

[14] Torvik V I, Smalheiser N R. Author name disambiguation in MEDLINE[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2009, 3(3): 1-29.

[15] Freeman R B, Huang W. Collaborating with people like me: Ethnic coauthorship within the United States[J]. Journal of Labor Economics, 2015, 33(S1): S289-S318.

[16] Myers K. The elasticity of science[J]. American Economic Journal: Applied Economics, 2020, 12(4): 103-134.

[17] Zhou S, Chai S, Freeman R B. Gender homophily: In-group citation preferences and the gender disadvantage[J]. Research Policy, 2024, 53(1): 104895.

[18] Krieger J L, Myers K R, Stern A D. How Important is Editorial Gatekeeping? Evidence from Top Biomedical Journals[J]. The Review of Economics and Statistics, 2021: 1-33.

[19] Carter C B, Blanford C F. All authors must now supply ORCID identifiers[J]. Journal of Materials Science, 2017, 52(11): 6147-6149.

[20] Moed H F, Aisati M, Plume A. Studying scientific migration in Scopus[J]. Scientometrics, 2013, 94: 929-942.

[21] Khurana P, Sharma K. Impact of h-index on author's rankings: an improvement to the h-index for lower-ranked authors[J]. Scientometrics, 2022, 127(8): 4483-4498.

[22] Zhao Z, Bu Y, Kang L, et al. An investigation of the relationship between scientists' mobility to/from China and their research performance[J]. Journal of Informetrics, 2020, 14(2): 101037.

[23] Aman V. Does the Scopus author ID suffice to track scientific international mobility? A case study based on Leibniz laureates[J]. Scientometrics, 2018, 117(2): 705-720.

[24] Kawashima H, Tomizawa H. Accuracy evaluation of Scopus Author ID based on the largest funding database in Japan[J]. Scientometrics, 2015, 103(3): 1061-1071.

[25] Boudry C, Durand-Barthez M. Use of author identifier services (ORCID, ResearcherID) and academic social networks (Academia. edu, ResearchGate) by the researchers of the University of Caen Normandy (France): A case study[J]. Plos one, 2020, 15(9): e0238583.

[26] Akers K G, Sarkozy A, Wu W, et al. ORCID author identifiers: A primer for librarians[J]. Medical Reference Services Quarterly, 2016, 35(2): 135-144.

[27] Web of Science [EB/OL].[2023-12-1].https://webofscience.help.clarivate.com/en-us/Content/wos-researcher-id.htm.

[28] AMiner [EB/OL].[2023-12-1].https://www.aminer.cn/introduction/.

[29] Song Y, Situ F, Zhu H, et al. To be the Prince to wake up Sleeping Beauty: The rediscovery of the delayed recognition studies[J]. Scientometrics, 2018, 117: 9-24.

[30] OpenAlex [EB/OL].[2023-12-1].https://openalex.org/about/.

[31] Priem J, Piwowar H, Orr R. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts[J]. arXiv preprint arXiv:2205.01833, 2022.

[32] Levin K, Cashore B, Bernstein S, et al. Overcoming the tragedy of super wicked problems: constraining our future selves to ameliorate global climate change[J]. Policy sciences, 2012, 45(2): 123-152.

[33] 刘玮辰,史冬波,李江.基于职业经历和引文网络的华人姓名消歧算法[J].信息资源管理学报,2020,10(06):82-89+100.(Liu Weichen, Shi Dongbo, Li Jiang. Name Disambiguation for Chinese Authors Using Their Career Experience and Citation Networks[J]. Journal of Information Resources Management, 2020,10(06):82-89+100.)

通讯作者（Corresponding author）:史冬波（Dongbo Shi），ORCID：0000-0003-1191-9103，Email：shidongbo@sjtu.edu.cn。

基金项目：本文系“国家自然科学基金面上项目”（项目编号：72374140）研究成果之一。The work is supported by “National Natural Science Foundation of China” (Grant No. 72374140).

作者贡献声明：

史冬波：提出研究思路和方案，数据收集、处理与分析，论文撰写和修改；

邓会：数据收集和校对，论文撰写和修改；

杨致简：数据收集和校对；

刘宁杰：数据处理；

刘余秀：数据收集；

毛宇飞：数据收集。

利益冲突声明：

所有作者声明不存在利益冲突关系。

支撑数据：

[1] 史冬波.科学家数据集. scientist.csv.

[2] 史冬波,邓会,杨致简,刘余秀,毛宇飞.数据库作者识别号及论文信息数据. (File) iddata.

附件

本文使用一批青年科学家的准确集数据，在 Web of Science, ORCID, Scopus, AMiner 和 OpenAlex 数据库中对这些科学家进行检索，并收集科学家在不同数据库中的标识号和论文数据。具体流程如下：第一步：在数据库检索界面输入科学家信息并进行检索，如姓名、工作机构等；第二步：根据检索结果与待测数据集中科学家的姓名、教育背景、工作机构、研究领域和起始时间等关键信息进行核对，以判定是否为同一位科学家。第三步：记录下科学家标识号，如果有多个与相关信息相匹配的科学家标识号，则记录下前三位最为匹配的科学家标识号，AMiner 记录科学家网页链接。第四步：确定对应科学家后，收集并保存相关论文信息至表格中，每个表格都以“author name_unique ID”命名，表格基本信息包括：论文 doi 号、标题、期刊、年份、入藏号、作者等。图 1 是此次进行检索的基本步骤。

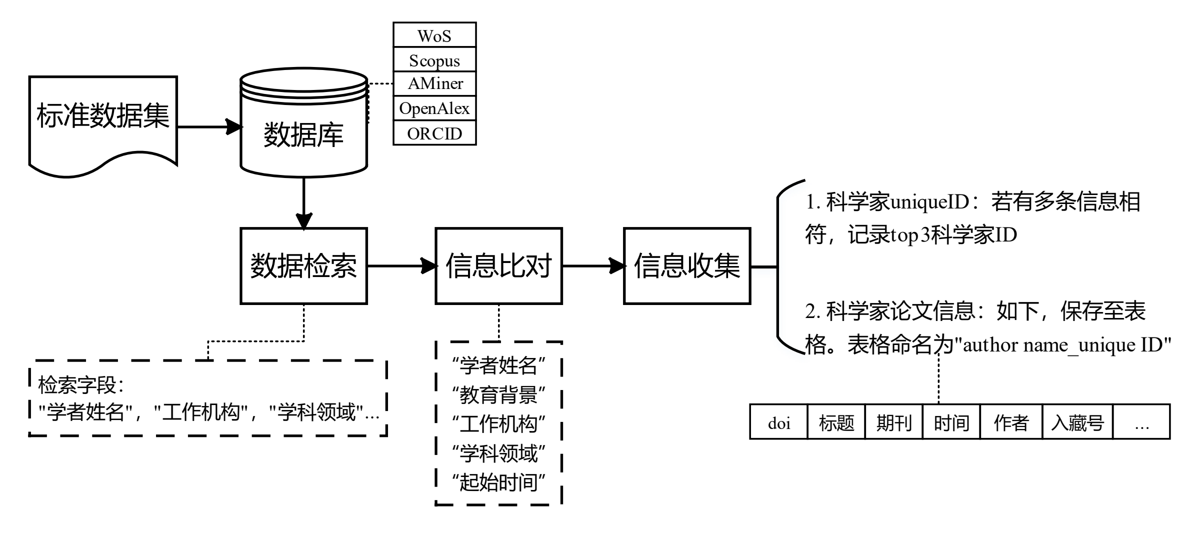


图 1 数据库检索基本流程图
Fig.1 Basic Flow Chart of Database Retrieval

以科学家 chen jianing 为例（图 2），记录在 5 个数据库中进行检索并保存科学家标识号和论文数据过程。

	A	B	C	D	E	F
1	uniqueID	inst	startyear	endyear	Familyname	Givenname
66	1_655	Dalian University of Technology	2003	2008	chen	jianing
67	1_655	Lund University	2009	2009	chen	jianing
68	1_655	CIC Nanogune & Donostia Internation	2010	2013	chen	jianing
69	1_655	University of the Basque Country	2010	2013	chen	jianing
70	1_655	Institute of Physics, CAS	2013	2021	chen	jianing

图 2 标准数据集中 chen jianing 基本信息
Fig.2 Basic information of chen jianing in the data set to be measured

1 Web of Science

第一步，进入网站检索界面，https://www.webofscience.com/wos/author/search；

第二步，输入科学家姓名，如图 3 所示；

第三步，根据科学家工作机构进行精炼（图 3）；

第四步，收集科学家的论文信息（如：doi，标题，期刊，年份，入藏号，作者，Researcher ID）至“2_chen jianing_HPU-2037-2023”。

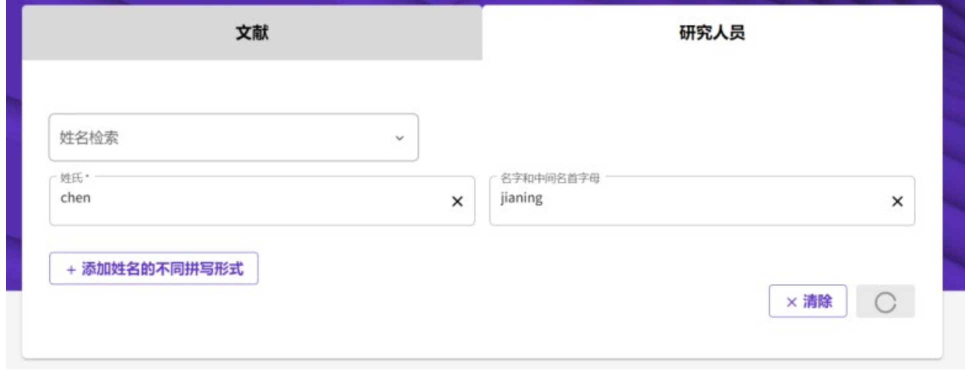


图 3 Web of Science 数据库中输入 chen jianing 姓名
Fig.3 Entering chen jianing's name in Web of Science database

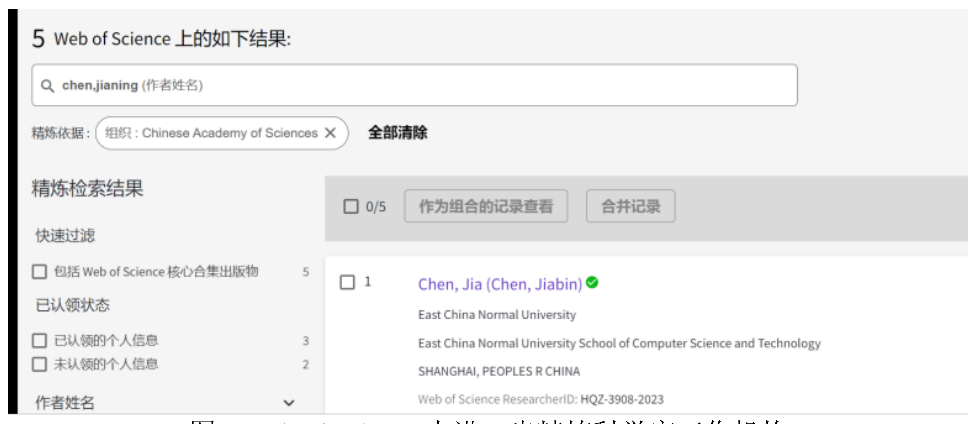


图 4 Web of Science 中进一步精炼科学家工作机构
Fig.4 Further refining the body of scientists working in Web of Science

2 ORCID

第一步，通过网址，进入官网 https://orcid.org；

第二步，输入相关信息：first name, last name, institution name，如图 6 所示；

第三步，核对关键信息是否一致（如 inst-startyear-endyear）(图 7)；

第四步，收集科学家的论文信息（如：doi，标题，期刊，年份，入藏号，作者，ORCID ID，Researcher ID）至“chen jianing_0000-0002-7525-1424”。

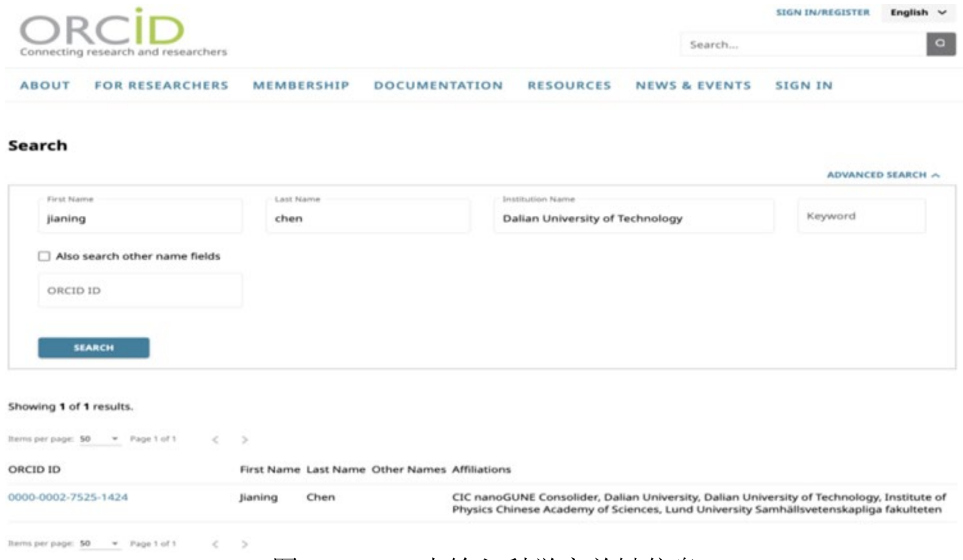


图 5 ORCID 中输入科学家关键信息

Fig.5 Key information for scientists entered in ORCID

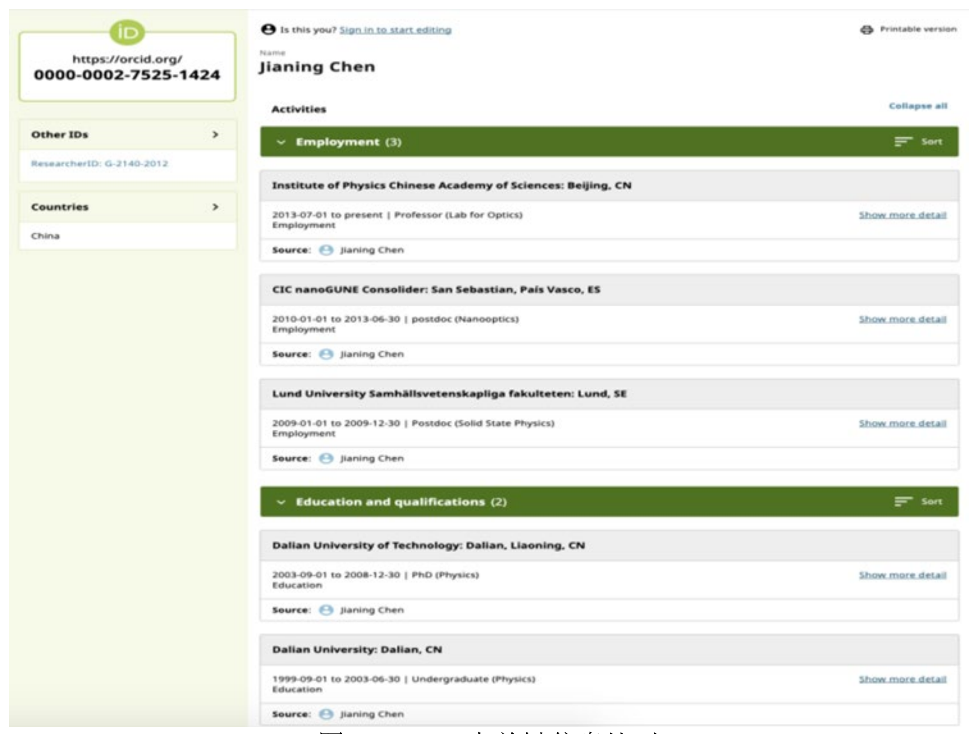


图 6 ORCID 中关键信息比对

Fig.6 Comparison of key information in ORCID

3 Scopus

第一步，进入网站进行检索，https://www.scopus.com/search/form.uri?display=authorLookup#author；

第二步，输入相关信息，如：first name, last name, institution name，并进行检索，如图 7；

第三步，核对教育经历、工作机构，学科背景、起始时间等是否一致，记录 ID 号，如（图 8）;

第四步，收集科学家的论文信息（如：doi，标题，期刊，年份，入藏号，作者，Scoups ID）至"Chen Jianing_55864209500.csv"。

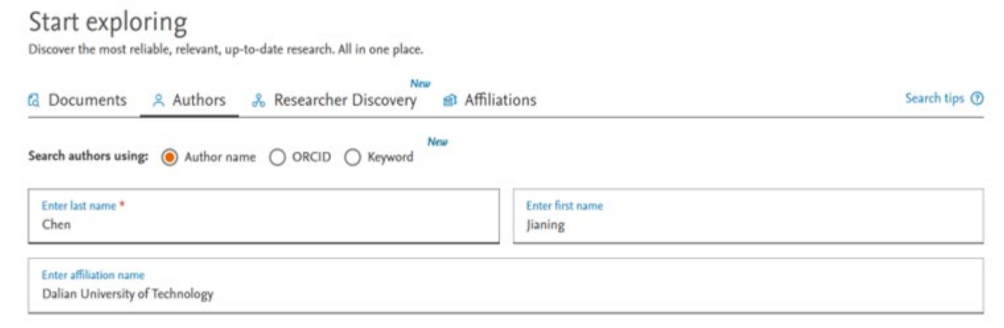


图 7 Scopus 中输入科学家关键信息

Fig.7 Entering key information about scientists in Scopus

Author information

Institution history

2008 - 2023	Institute of Physics Chinese Academy of Sciences
2008 - 2023	Chinese Academy of Sciences
2017 - 2018	University of Chinese Academy of Sciences
2018	Collaborative Innovation Center of Quantum Matter
2015 - 2017	Collaborative Innovation Center of Quantum Matter
2017	Collaborative Innovation Center of Quantum Matter
2017	Collaborative Innovation Center of Quantum Matter
2011 - 2014	CIC nanoGUNE
2011 - 2013	Donostia International Physics Center
2011 - 2013	CSIC UPV Centro de Fisica de Materials CFM
2010	NanoLund, Lund University
2010	Lunds Universitet
2008	Dalian University of Technology

图 8 Scopus 中科学家关键信息比对

Fig.8 Comparison of scientists' key information in Scopus

4 AMiner

第一步，进入网站，https://www.AMiner.org/;

第二步，输入相关信息，如：first name, last name, institution 等，进行检索，如图 9;

第三步，查看科学家基本信息：教育背景、工作机构、起始时间等，判断关键信息是否一致（图 10）；

第四步，收集科学家的论文信息（如：doi，标题，期刊，年份，作者）至“chen jianing_AMiner.cn/profile/jianing-chen/54056041dabfae91d3fdb590”。

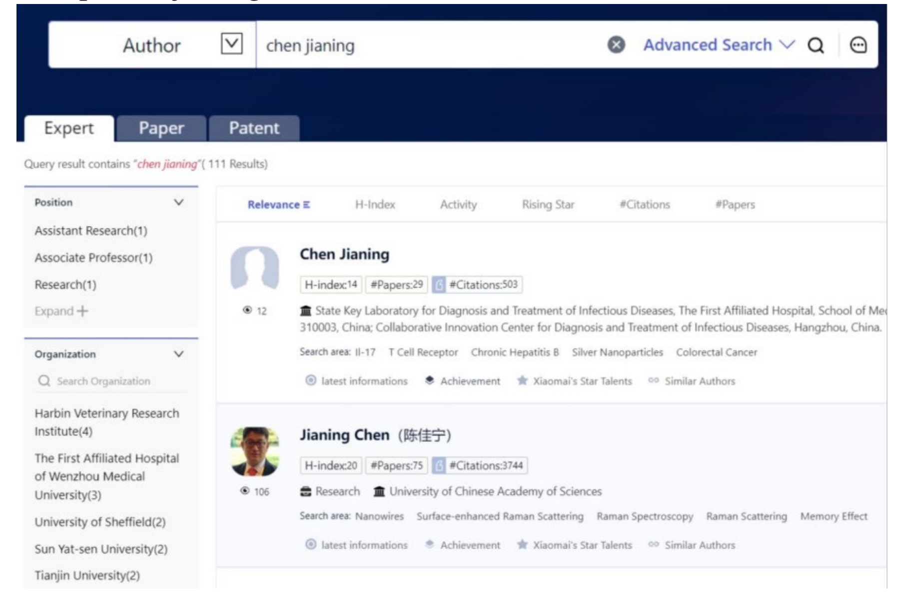


图 9 AMiner 中科学家 chen jianing 关键信息

Fig.9 Key information about scientist chen jianing in AMiner



图 10 AMiner 中输入科学家姓名 chen jianing

Fig.10 Entering the scientist's name chen jianing in AMiner

在 Web of Science, ORCID, Scopus, AMiner, and OpenAlex 数据库中收集科学家 chen jianing 论文数据并保存至表格中（图 11），表格以“author name_unique ID”命名，AMiner 科学家 ID 为科学家网页链接。

2_chen jianing_HPU-2037-2023	2023/11/4 21:36	XLS 工作表	31 KB
chen jianing_0000-0002-7525-1424	2023/11/4 21:36	XLS 工作表	5 KB
Chen Jianing_55864209500	2023/11/4 21:36	XLS 工作表	12 KB
chenjianing_aminer.cnprofilejianing-chen54...	2023/11/21 0:20	XLS 工作表	9 KB

图 11 WoS, ORCID, Scopus 和 AMiner 数据库中 chen jianing 的论文数据表格
Fig.11 Data table of chen jianing's papers in WoS, ORCID, Scopus and AMiner databases